

BAYESIAN SELECTION OF REGRESSION MODEL FOR PROBABILISTIC CHARACTERIZATION OF ROCK UNIAXIAL COMPRESSIVE STRENGTH

Adeyemi E. Aladejare¹ and Yu Wang¹

¹ Department of Architecture and Civil Engineering, Shenzhen Research Institute, City University of
Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
e-mail: aaladejar2-c@my.cityu.edu.hk

Keywords: Regression Model, Occurrence Probability, Site/Deposit, Bayesian Framework, Probabilistic Characterization, Markov Chain Monte Carlo Simulation.

Abstract. *Rock engineers and practitioners use regression equations/models between uniaxial compressive strength, UCS and point load index, $Is_{(50)}$ to estimate UCS when there is difficulty in direct determination of UCS from laboratory tests. $Is_{(50)}$ has been reported as an indirect measure of the compressive strength of rocks. However, there are many equations in the literature relating $Is_{(50)}$ to UCS for all rock types. Selecting the appropriate model for UCS estimation in a particular site/deposit becomes problematic. This is because UCS of rocks, like other geomechanical properties are products of different geological processes that rocks are subjected to, which makes them to be variable even within a formation or deposit. This study presents an approach for selecting regression equation for estimating UCS from $Is_{(50)}$, using only a limited number of $Is_{(50)}$ data obtained from a specific deposit or site. The approach works by comparing the occurrence probability of each model for a given set of observation (i.e. $Is_{(50)}$) data obtained from a site/deposit. The most appropriate model is the model with the highest occurrence probability for the given set of observation data. This approach is different from previous works that need both UCS and $Is_{(50)}$ data to draw comparison; instead it selects the appropriate model using the $Is_{(50)}$ data only. This is important because the time when there is need for selection and use of regression model is when the UCS data are not available. The selected model is then used in a Bayesian framework to integrate the prior knowledge about UCS with the limited number of site-specific $Is_{(50)}$ data available for probabilistic characterization of UCS. This is achieved by using Markov Chain Monte Carlo (MCMC) simulation to generate UCS samples from resulting posterior PDF of the Bayesian framework. Statistical analyses are then performed on the UCS samples to obtain its mean, standard deviation and full probabilistic distribution.*

1 INTRODUCTION

Mechanical behaviour of rocks is required for safe design of structures in civil, mining and petroleum engineering applications. Uniaxial compressive strength (UCS) of rock is a mechanical property of rock that is very important during geotechnical characterization because of its role in the design and analysis of geotechnical problems [1]. Rock engineers use UCS more than other rock properties, according to a survey reported by Bieniaskwi [2]. Hoek [3] emphasized that UCS is a useful parameter when considering a variety of issues encountered during blasting, excavation, and supporting in engineering works. The UCS of rock can be determined in the laboratory by uniaxial compression test. However, carrying out compression test may not be possible at all times, this is because it is not always possible to extract proper cores for sampling purpose in highly weathered rocks or thinly bedded or densely fractured and also when there are time and resources constraints [4, 5]. When the direct determination of UCS is not possible, there is need to estimate UCS from other strength index widely used by rock engineers. The point load strength ($Is_{(50)}$) has been found to be an efficient index in estimating UCS [6, 7]. Broch and Franklin [6] explained that the need for a relationship between $Is_{(50)}$ and UCS dated back to the period after development of the point loading method. Generally, the point load test can be applied axially or diametrically to rock cores or irregular lumps. The required testing apparatus is light and usually portable, and more importantly, the test are inexpensive. However, different studies have shown that there is no single regression equation relating $Is_{(50)}$ to UCS for all rock types (e.g., [6-8]). The database of the empirical equations between UCS and $Is_{(50)}$ has been reported in Wang and Aladejare [9]. With many regression models between UCS and $Is_{(50)}$ available in the literature even for the same rock type, there is need to have a rational method for selecting appropriate model for estimating UCS in a specific site/deposit. Model selection for estimating UCS from $Is_{(50)}$ is mostly needed when there is no UCS at all [9]. This makes it important to develop a method for selecting model using just $Is_{(50)}$ data available from the site.

This paper presents a method for site-specific selection of regression model using just the limited $Is_{(50)}$ from the site/deposit. The method integrates the limited site $Is_{(50)}$ data with the prior knowledge of UCS under Bayesian framework to select the appropriate model and the selected model is further used in probabilistic characterization of UCS to obtain its statistics and full distribution.

2 SITE-SPECIFIC SELECTION OF REGRESSION MODEL

Site-specific selection of regression model can be achieved by using just limited $Is_{(50)}$ data available from the site. This is done by comparing the occurrence probability $P(M_j|Data)$ of different models M_j given observed $Is_{(50)}$ data from the site (e.g., [9-11]). The model with the maximum occurrence probability is the most appropriate for such site. Using Bayes' theorem, $P(M_j|Data)$ can be obtained as expressed in Eq. (1).

$$P(M_j|Data) = P(Data|M_j)P(M_j)/P(Data) \quad j = 1, 2, \dots, k \quad (1)$$

In Eq. (1), $Data$ denote site-specific $Is_{(50)}$ data, M_j is the candidate model, $P(Data | M_j)$ is the probability of observing $Data$ given a candidate model M_j (e.g., [9-11]). $P(M_j)$ is the prior probability of M_j which reflects the prior knowledge of M_j . The prior probability of each model is taken as $1/k$ in the absence of any prevailing prior knowledge about the models, while $P(Data)$ is a normalizing constant.

The theorem of total probability can be used to obtain $P(Data | M_j)$ as expressed in Eq. (2).

$$P(Data|M_j) = \iint P(Data|\mu, \sigma, M_j) \times P(\mu, \sigma|M_j) d\mu d\sigma \quad (2)$$

In Eq. (2), $P(Data|\mu, \sigma, M_j)$ is the likelihood function, which is the joint conditional probability density function, PDF of $Data$ for given M_j and a set of model parameters (e.g. mean μ and standard σ of UCS). $P(\mu, \sigma | M_j)$ is the prior distribution of μ and σ of UCS.

In formulating the likelihood function in Eq. (2), variability of UCS and the transformation uncertainty of using regression model for estimating UCS from $Is_{(50)}$ data are considered. UCS is taken to follow a normal distribution with a mean μ and standard deviation σ , expressed in Eq. (3).

$$UCS = \mu + \sigma z \quad (3)$$

where z is a standard Gaussian random variable.

If the format of regression model for estimating UCS from $Is_{(50)}$ is as expressed in Eq. (4).

$$UCS = aIs_{(50)} \quad (4)$$

where a is the conversion factor from $Is_{(50)}$ to UCS. The general transformation model for each model is as expressed in Eq. (5).

$$Is_{(50)} = bUCS + \varepsilon \quad (5)$$

where b is the inverse of a in Eq. (4) and ε , which is the modelling error is a Gaussian random variable with zero mean and standard deviation σ_ε .

The general likelihood model is obtained by combining Eq. (3) and Eq. (5) and expressed in Eq. (6).

$$Is_{(50)} = b(\mu + \sigma z) + \varepsilon \quad (6)$$

if the inherent variability is independent of the transformation uncertainty (i.e. z is independent of ε), $Is_{(50)}$ is a Gaussian random variable with a mean of $(b\mu)$ and standard deviation of $\sqrt{(b\sigma)^2 + \sigma_\varepsilon^2}$. The site-specific point load data (i.e., $Data = \{Is_{(50)_s}, s = 1, 2, \dots, n_t\}$) can be considered as n_t independent realizations of the Gaussian random variable $Is_{(50)}$. The likelihood function $P(Data|\mu, \sigma, M_j)$ for the site-specific $Is_{(50)}$ is as expressed in Eq. (7).

$$P(Data|\mu, \sigma, M_j) = \prod_{s=1}^{n_t} \frac{1}{\sqrt{2\pi}\sqrt{(b\sigma)^2 + \sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2}\left[\frac{Is_{(50)_s} - (b\mu)}{\sqrt{(b\sigma)^2 + \sigma_\varepsilon^2}}\right]^2\right\} \quad (7)$$

The prior distribution $P(\mu, \sigma | M_j)$ in Eq. (2) can be taken as joint uniform distribution of μ and σ , with respective minimum values of μ_{min} and σ_{min} and respective maximum values of μ_{max} and σ_{max} . This type of distribution is mostly assumed when the prior knowledge is relatively uninformative. $P(\mu, \sigma | M_j)$ is as expressed in Eq. (8).

$$P(\mu, \sigma | M_j) = \begin{cases} \frac{1}{\mu_{max} - \mu_{min}} \times \frac{1}{\sigma_{max} - \sigma_{min}} & \text{for } \mu \in [\mu_{min}, \mu_{max}] \text{ and } \sigma \in [\sigma_{min}, \sigma_{max}] \\ 0 & \text{others} \end{cases} \quad (8)$$

However, the approach presented in this paper is general and can also be used for relatively informative prior knowledge like an arbitrary histogram type, normal distribution etc. The Bayesian framework for probabilistic characterization of UCS using the selected model from model selection is presented in the next section.

3 PROBABILISTIC CHARACTERIZATION OF UCS

The uniaxial compressive strength is modeled by a random variable UCS which follows a normal distribution with a mean μ and standard deviation σ . For a given set of prior knowledge and site-specific $Is_{(50)}$ data, there are many sets of possible values of μ and σ . For a given set of prior knowledge and project-specific $Is_{(50)}$ data, there are many sets of possible values of μ and σ . Each set of μ and σ has its corresponding occurrence probability, which is defined by a joint conditional PDF, $P(\mu, \sigma | \text{Data})$. Using the Theorem of Total Probability, the PDF of the UCS for a given set of prior knowledge and site-specific $Is_{(50)}$ data is expressed in Eq. (9).

$$P(\text{UCS} | \text{Data}) = \iint P(\text{UCS} | \mu, \sigma) \times P(\mu, \sigma | \text{Data}) d\mu d\sigma \quad (9)$$

The joint conditional PDF $P(\mu, \sigma | \text{Data})$ can be simplified using Bayesian framework as expressed in Eq. (10).

$$P(\mu, \sigma | \text{Data}) = KP(\text{Data} | \mu, \sigma)P(\mu, \sigma) \quad (10)$$

where K is a normalizing constant. $P(\text{Data} | \mu, \sigma)$ is the likelihood function of the selected model, and $P(\mu, \sigma)$ is the prior distribution of μ and σ given in Eqs. (7) and (8) respectively.

Eqs. (9) and (10) can be combined to give Eq. (11).

$$P(\text{UCS} | \text{Data}) = K \iint P(\text{UCS} | \mu, \sigma)P(\text{Data} | \mu, \sigma)P(\mu, \sigma) d\mu d\sigma \quad (11)$$

Bayesian equivalent approach developed by Wang and Cao [10] is used to transform the posterior PDF $P(\text{UCS} | \text{Data})$ into large number of UCS samples. The approach integrates Markov Chain Monte Carlo (MCMC) simulation with the Bayesian method to generate samples of UCS. Their approach is effective in eliminating the difficulty in generating samples from an arbitrary and complicated posterior PDF. Conventional statistics are then performed on the samples to determine the mean and standard deviation of UCS. The PDF and cumulative distribution function, CDF, of UCS are also constructed.

4 ILLUSTRATIVE EXAMPLE

The proposed Bayesian selection approach is applied to select the most appropriate model for granite at Malanjkhanda Copper project site in the state of Madhya Pradesh, India using the nineteen $Is_{(50)}$ values reported by Mishra and Basu [12]. Their data also includes twenty direct measurements of UCS values which will be used for comparing and validating the result of the proposed approaches with direct measurement data in hand. The nineteen $Is_{(50)}$ and twenty UCS data are shown in columns two and three of Table 1 respectively.

Specimen No	Is ₍₅₀₎ (MPa)	UCS (MPa)
1	8.35	139.04
2	10.85	177.37
3	10.02	167.17
4	9.92	176.75
5	11.73	160.82
6	14.13	198.15
7	10.63	148.34
8	6.93	117.95
9	8.49	134.76
10	7.87	124.89
11	8.41	138.22
12	7.85	130.06
13	5.99	122.74
14	Invalid	201.73
15	7.29	153.55
16	11.36	182.33
17	9.23	150.42
18	6.92	127.47
19	9.72	158.69
20	5.66	91.48

Table 1: Laboratory test results of granite from Malanjkhanda Copper project, India [12]

To perform site-specific selection of regression model and probabilistic characterization of UCS, four models are used for illustration. The model by Broch and Franklin [6] is given in Eq. (12) and taken as model 1, M_1 .

$$UCS = 23.7Is_{(50)} \quad (12)$$

Tugrul and Zarif [13] developed a regression given in Eq. (13) and it is taken as model 2, M_2 .

$$UCS = 15.25Is_{(50)} \quad (13)$$

The model from the least square regression analysis of $Is_{(50)}$ and UCS data points reported by Ghosh and Srivastava [14] is given in Eq. (14) and taken as model 3, M_3 .

$$UCS = 13.85Is_{(50)} \quad (14)$$

Chau and Wong [7] developed a regression model given in Eq. (15).

$$UCS = 16.31Is_{(50)} \quad (15)$$

The regression model is taken as model 4, M_4 , in this study. The constant factor of 16.31 is obtained through the re-analysis of their 21 data points.

The summary of the parameters of each model used in this study is presented in Table 2.

Parameter	M_1	M_2	M_3	M_4
b	0.042	0.066	0.072	0.061
σ_ε	1.500	0.292	1.424	2.073

Table 2: Parameters of the models compared

σ_e is calculated using the original data reported by the authors.

Kulhawy [15] reported the ranges of UCS values for igneous rock as from 3.65 MPa to 355 MPa. The range of $Is_{(50)}$ for granite compiled in the literature is from 1.04MPa to 14.85 MPa (e.g., [7, 13]). This range is combined with the range of site $Is_{(50)}$, to estimate the likely range of UCS at this site by linear interpolation, since the relation between UCS and $Is_{(50)}$ is linear (see Eqs. (12-15)). This gives a UCS range from 121 MPa to 337 MPa. Using six-sigma rule, the maximum standard deviation of UCS is 36.0 MPa and the minimum standard deviation of UCS is taken as 0 MPa. This set of ranges (i.e. μ [121 MPa, 337 MPa] and σ [0 MPa, 36 MPa]) is taken as prior knowledge of the site. The nineteen $Is_{(50)}$ data points are then used in the model selection approach and probabilistic characterization of UCS. The results are discussed in the next subsections.

4.1 Most appropriate site-specific regression model

The result of the calculated evidence and occurrence probability for the four models investigated is shown in Table 3.

Regression model M_j	Evidence $P(\text{Data} M_j)$	Occurrence probability $P(M_j \text{Data})$
M_1	8.3247E-20	0.312
M_2	3.3537E-20	0.126
M_3	4.1654E-20	0.156
M_4	1.0834E-19	0.406

Table 3: Comparison of the evidence of models

From the result in Table 3, Model 4 is the most appropriate model for the site because it has the highest occurrence probability. In addition, the regression lines of the models and the site-specific data are plotted in Figure 1.

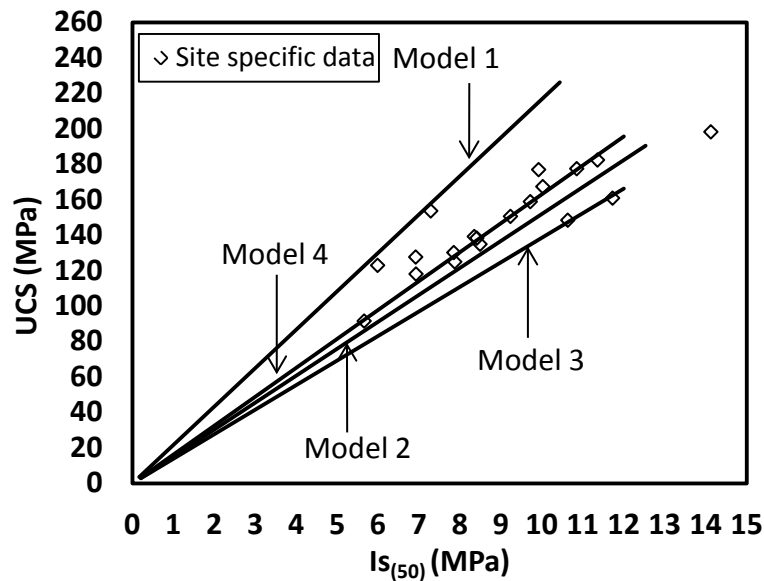


Figure 1: Plot of the site-specific UCS and $Is_{(50)}$ data with the regression lines of the models

Figure 1 shows that the site-specific $Is_{(50)}$ and UCS data points plots more closely to the regression lines of model 4, M_4 than that of model 1, M_1 , model 2, M_2 and model 3, M_3 . This observation agrees with the result of the evidence calculation using only the $Is_{(50)}$ data points, with model 4, M_4 having the highest value of evidence. The proposed approach selected the most appropriate model for this specific site in a transparent and rational way. Note that most times, there is no direct measurement of UCS data to plot this relationship between UCS and $Is_{(50)}$ data. The proposed approach is a useful tool for selection of the most appropriate site-specific regression model using just limited number of $Is_{(50)}$ data.

4.2 Probabilistic characterization of UCS

The most appropriate model is combined with the nineteen site $Is_{(50)}$ data in a Bayesian equivalent sample approach to generate 30,000 equivalent UCS samples. The samples are simulated using a two-dimensional grid over the space of μ and σ with an interval of 0.5 MPa in both directions. The scatter plot of the samples is presented in Figure 2.

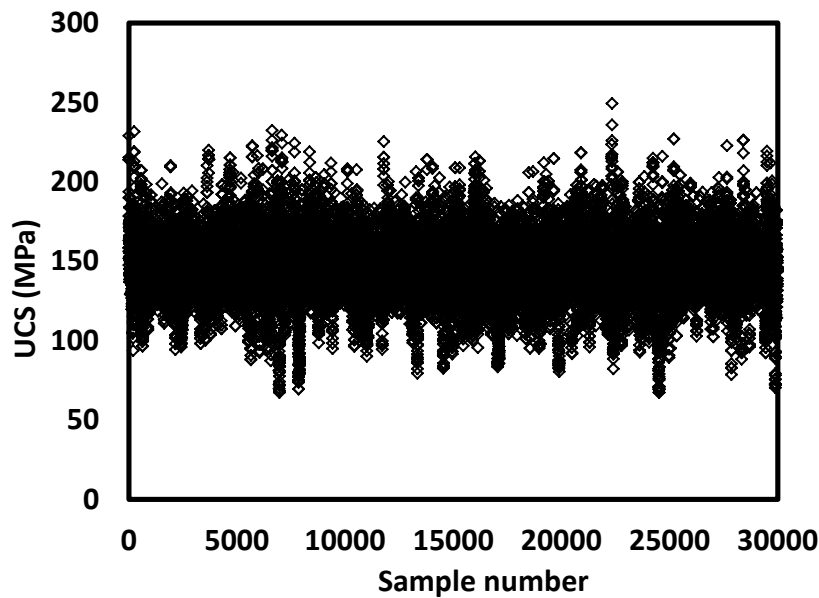


Figure 2: Scatter plot of the Bayesian equivalent samples of UCS

The statistics of the generated UCS samples are presented in Table 4.

Approaches	Bayesian Equivalent Sample	Compression Test (MPa)	Relative Difference (%)
Mean (MPa)	147.8	150.1	1.5
Standard Deviation (MPa)	18.7	28.3	34.0

Table 4: Summary of the statistics of UCS

Table 4 shows that the mean and standard deviation of the generated UCS samples is 147.8 MPa and 18.7 MPa respectively. Comparing the mean and standard deviation values of simulated samples with that from compression test which are also included in Table 4, the relative difference in mean values is 1.5 %, while the relative difference in standard deviation is 34.0 %. The small difference between the mean values indicates that the mean agrees well with

each other, suggesting that the approaches perform satisfactorily. The difference in standard deviation between the direct measurement and Bayesian equivalent sample approach is relatively large. This might be as a result of the limited number of UCS data from the direct measurement which may not be sufficient to provide a reasonable estimate of UCS standard deviation at this particular site.

Figure 3 shows the PDF of the 30,000 equivalent UCS samples plotted by solid line and the 20 UCS values from compression test plotted with open triangles. Figure 4 shows the UCS CDFs estimated from the cumulative frequency diagrams of the 30,000 generated samples plotted by solid line and the 20 direct measurement results plotted by open triangles.

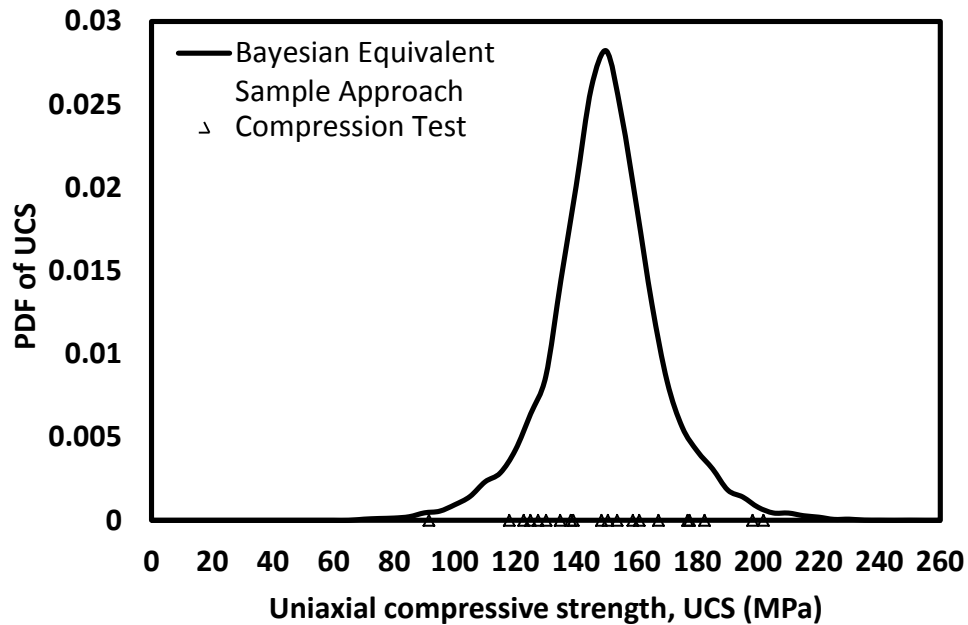


Figure 3: Probability density function (PDF) for uniaxial compressive strength

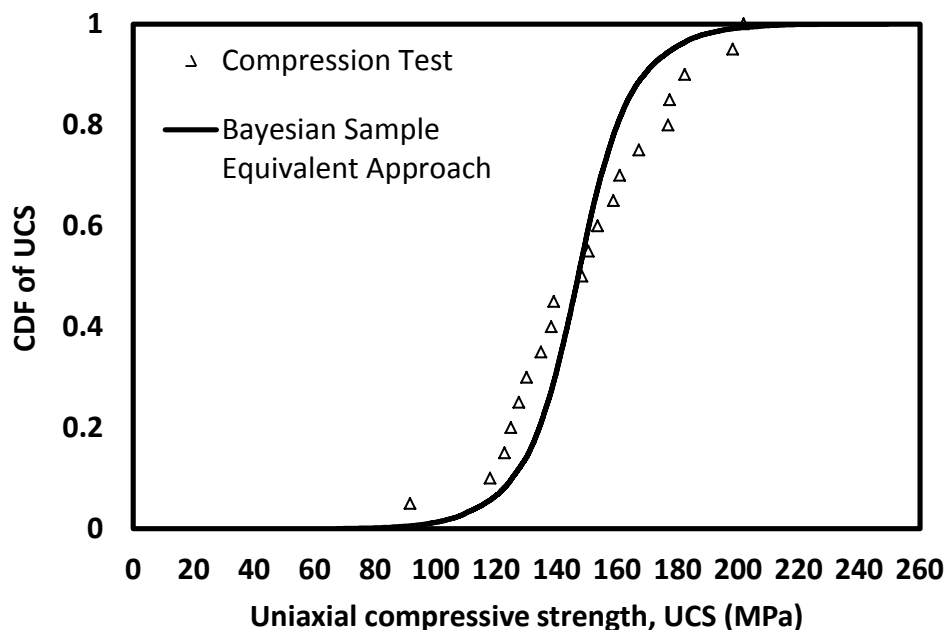


Figure 4: Validation of the probability distribution of the UCS from Bayesian equivalent samples

The PDF of the generated UCS samples are well within the range of UCS samples values from direct measurements. Also, there is good agreement between the CDFs of UCS from the direct measurement by compression tests and that of the Bayesian equivalent sample approach. These agreements showed that the information contained in the UCS samples from Bayesian equivalent sample approach is consistent with the one obtained from the direct measurement of UCS by compression tests. This probabilistic characterization of rock UCS usually require a large amount of data from in-situ and/or laboratory tests, which could be tedious, time and energy consuming or even impossible because of the strength nature of some rocks.

5 CONCLUSIONS

The proposed approach presented in this study is used to select the most appropriate model for specific site by integrating the information from the regression models considered and site-specific $Is_{(50)}$ data. The most appropriate regression model selected is further used in the development of the Bayesian equivalent sample approach for characterization of UCS, by combining it with the prior knowledge and the nineteen site-specific $Is_{(50)}$ data available. Bayesian equivalent sample approach transformed the updated information from the prior knowledge and nineteen site-specific $Is_{(50)}$ data into a large number of UCS samples. Note that, given a limited site-specific $Is_{(50)}$ data, this approach generates a large number of UCS data points that is sufficient enough for conventional statistical analysis. This approach is useful during site investigations into project that requires rock properties for subsequent applications in engineering design and construction, also in tunneling, mining and other projects involving the use of rock during construction. In addition, the samples from this approach can be used directly in Monte Carlo simulation-based reliability analysis and designs [e.g., 16].

ACKNOWLEDGEMENT

The work described in this paper was supported by a grant from the National Natural Science Foundation of China (Project No. 51208446). The financial support is gratefully acknowledged.

REFERENCES

- [1] R. Nazir, E. Momeni, D.J. Armaghani, M.F.M. Amin, Correlation between unconfined compressive strength and indirect tensile strength of limestone rock. *Electronic Journal of Geotechnical Engineering*, **18**, 1737-1746, 2013.
- [2] Z.T. Bieniawski, Estimating the strength of rock materials. *Journal of the Southern African Institute of Mining and Metallurgy*, **74**, 312-320, 1976.
- [3] E. Hoek, Rock mechanics laboratory testing in the context of a consulting engineering organization. *International Journal of Rock Mechanics and Mining Science & Geomechanics Abstract*, **14**, 93– 101, 1977.
- [4] S. Kahraman, Evaluation of simple methods for assessing the uniaxial compressive strength of rock. *International Journal of Rock Mechanics and Mining Science*, **38**, 981–994, 2001.

- [5] N. Yesiloglu-Gultekin, C. Gokceoglu, E.A. Sezer, Prediction of uniaxial compressive strength of granitic rocks by various nonlinear tools and comparison of their performances. *International Journal of Rock Mechanics and Mining Science*, **62**, 113–122, 2013.
- [6] E. Broch, J.A. Franklin, The point-load strength test. *International Journal of Rock Mechanics and Mining Science*, **28**, 669–697, 1972.
- [7] K.T. Chau, R.H.C. Wong, Uniaxial compressive strength and point load strength of rocks. *International Journal of Rock Mechanics and Mining Science*, **33**, 183–188, 1996.
- [8] A. Basu, A. Aydin, Predicting uniaxial compressive strength by point load test: significance of cone penetration. *Rock Mechanics and Rock Engineering*, **39**, 483–490, 2006.
- [9] Y. Wang, A.E. Aladejare, Selection of site-specific regression model for characterization of uniaxial compressive strength of rock. *International Journal of Rock Mechanics and Mining Science*, **75**, 73–81, 2015.
- [10] Y. Wang, Z. Cao, Probabilistic characterization of Young's modulus of soil using equivalent samples. *Engineering Geology*, **159**, 106–118, 2013.
- [11] Z. Cao, Y. Wang, Bayesian model comparison and characterization of undrained shear strength. *Journal of Geotechnical and Geoenvironmental Engineering*, **140**(6), 04014018, 2014.
- [12] D.A. Mishra, A. Basu, Use of the block punch test to predict the compressive and tensile strengths of rocks. *International Journal of Rock Mechanics and Mining Science*, **51**, 119–127, 2012.
- [13] A. Tugrul, I.H. Zarif, Correlation of mineralogical and textural characteristics with engineering properties of selected granitic rocks from Turkey. *Engineering Geology*, **51**, 303–317, 1999.
- [14] D.K. Ghosh, M. Srivastava, Point-load strength: an index for classification of rock material. *Bulletin of Engineering Geology and the Environment*, **44**, 27–33, 1991.
- [15] F.H. Kulhawy, Rock deformation properties of rock and rock discontinuities. *Engineering Geology*, **9**, 327–350, 1975.
- [16] Y. Wang, MCS-based probabilistic design of embedded sheet pile walls. *Georisk*, **7**(3), 151–162, 2013.